A Probabilistic Quantitative Analysis of Probabilistic-Write/Copy-Select

Christel Baier[1], Benjamin Engel[2], Sascha Klüppelholz[1], Steffen Märcker[1], Hendrik Tews[2], Marcus Völp[2]

[1]Institute for Theoretical Computer Science

[2]Operating-Systems Group

Technische Universität Dresden, Germany

NASA Formal Methods Symposium (NFM'13)
May 16, 2013

Motivation

Observation: traditional locking does not scale any more

- atomic operations are slow and become increasingly expensive
- locking schemes will become more complex and
- scalability becomes problematic on future hardware systems

Motivation

Observation: traditional locking does not scale any more

- atomic operations are slow and become increasingly expensive
- locking schemes will become more complex and
- scalability becomes problematic on future hardware systems

Idea: Probabilistic-Write/Copy-Select (PWCS) [Mc Guire'11]

- no locks, no atomic operations
- make inconsistencies detectable (e.g., tags, hashes)
- sufficiently high probability to find a consistent replica

Motivation

Observation: traditional locking does not scale any more

- atomic operations are slow and become increasingly expensive
- locking schemes will become more complex and
- scalability becomes problematic on future hardware systems

Idea: Probabilistic-Write/Copy-Select (PWCS) [Mc Guire'11]

- no locks, no atomic operations
- make inconsistencies detectable (e.g., tags, hashes)
- sufficiently high probability to find a consistent replica

Properties of PWCS

- measure-based experiments [Mc Guire'11]: promising approach
- promising to work with more relaxed memory models
- instance of a new class of algorithms (inherent randomness)

The PWCS protocol [Mc Guire'11]

Writer

```
for i=1..n
  r = replica[i];
  r.end_tag++;
  r.write_data();
  r.begin_tag++;
endfor
```

Replica

B_1	$Data_1$	E_1
B_2	$Data_2$	E_2
$\vdots$	$\vdots$	$\vdots$
B_n	$Data_n$	E_n

Reader

```
for i=n..1
  r = replica[i];
  ta = r.begin_tag;
  r.copy_data();
  tb = r.end_tag;
  if (ta == tb)
   return data;
endfor

// error case
```

The PWCS protocol [Mc Guire'11]

Writer

```
for i=1..n
  r = replica[i];
  r.end_tag++;
  r.write_data();
  r.begin_tag++;
endfor
```

CTMC model

Replica

B_1	$Data_1$	E_1
B_2	$Data_2$	E_2
	$\vdots$	
B_n	$Data_n$	E_n

transition system model

Reader

```
for i=n..1
  r = replica[i];
  ta = r.begin_tag;
  r.copy_data();
  tb = r.end_tag;
  if (ta == tb)
   return data;
endfor

// error case
```

CTMC model

Contribution (NFM'13)

- continuous-time Markov chain (CTMC) model for PWCS with multiple writers
- identify quantitative measures for the evaluation of PWCS
- formalization of quantitative measures in terms of continuous stochastic reward logic (CSRL)
- formal quantitative analysis of PWCS using the probabilistic model checker PRISM

Outline

Continuous-time Markov chain (CTMC)

Definition (CTMC)

A CTMC is a tuple $\mathcal{M} = \langle S, Act, R, \mu \rangle$, where

- S a finite state space,
- Act a finite set of action names,
- $R : S \times Act \times S \to \mathbb{R}_{\geq 0}$ the rate matrix of $\mathcal{M}$,
- $\mu : S \to [0,1]$ a distribution on S, i.e., $\sum_{s \in S} \mu(s) = 1$

Continuous-time Markov chain (CTMC)

Definition (CTMC)

A CTMC is a tuple $\mathcal{M} = \langle S, Act, R, \mu \rangle$, where

- S a finite state space,
- Act a finite set of action names,
- $R : S \times Act \times S \rightarrow \mathbb{R}_{\geq 0}$ the rate matrix of $\mathcal{M}$,
- $\mu : S \rightarrow [0, 1]$ a distribution on S, i.e., $\sum_{s \in S} \mu(s) = 1$

Probability for $s \xrightarrow{\lambda:\alpha} s'$ ready to fire in $[0, t]$ is

$$1 - e^{-\lambda t}$$

Thus, the average delay of this transition is $1/\lambda$.

Continuous-time Markov chain (CTMC)

Definition (CTMC)

A CTMC is a tuple $\mathcal{M} = \langle S, Act, R, \mu \rangle$, where

- S a finite state space,
- Act a finite set of action names,
- $R : S \times Act \times S \to \mathbb{R}_{\geq 0}$ the rate matrix of $\mathcal{M}$,
- $\mu : S \to [0, 1]$ a distribution on S, i.e., $\sum_{s \in S} \mu(s) = 1$

Probability for firing $s \xrightarrow{\lambda:\alpha} s'$ in $[0, t]$ is

$$P(s, \alpha, s') \cdot \left(1 - e^{-E(s) \cdot t}\right)$$

where $E(s)$ denotes the exit rate of state s, i.e., the sum of the rates of all outgoing transitions of state s.

Continuous-time Markov chain (CTMC)

Definition (CTMC)

A CTMC is a tuple $\mathcal{M} = \langle S, Act, R, \mu \rangle$, where

- S a finite state space,
- Act a finite set of action names,
- $R : S \times Act \times S \to \mathbb{R}_{\geq 0}$ the rate matrix of $\mathcal{M}$,
- $\mu : S \to [0, 1]$ a distribution on S, i.e., $\sum_{s \in S} \mu(s) = 1$

Probability for firing $s \xrightarrow{\lambda:\alpha} s'$ in $[0, t]$ is

$$\lambda / E(s) \cdot \left(1 - e^{-E(s) \cdot t}\right)$$

where $E(s)$ denotes the exit rate of state s, i.e., the sum of the rates of all outgoing transitions of state s.

PWCS composed CTMC model

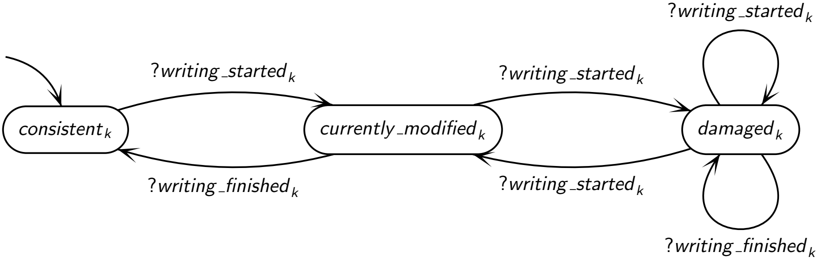
Product of CTMC for the writers, CTMC for the readers, and ordinary (non-stochastic) transition systems for the replicas.

$$\frac{s \xrightarrow{\lambda:\alpha} s'}{\langle s, \overline{x} \rangle \xrightarrow{\lambda:\alpha} \langle s', \overline{x} \rangle} \qquad \frac{w \xrightarrow{\lambda:!a} w', \quad r \xrightarrow{?a} r'}{\langle w, r, \overline{y} \rangle \xrightarrow{\lambda:a} \langle w', r', \overline{y} \rangle}$$

$\overline{x}$: local states of all other components

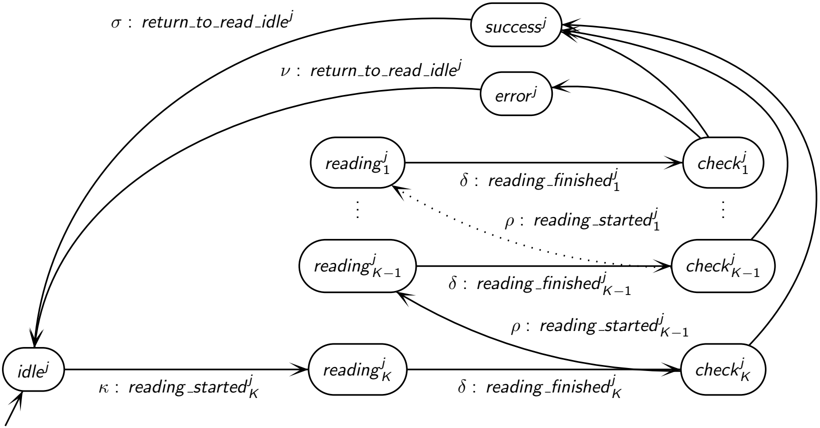
$\overline{y}$: local states of all readers and remaining writers and replicas

PWCS model



Transition system model of a replica

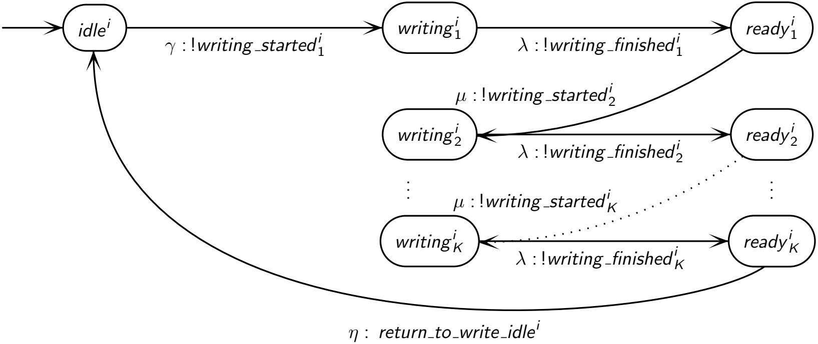
PWCS model



CTMC model of a reader

PWCS model

CTMC model of a writer



Outline

Interesting quantitative measures

M1: probability to successfully read the data

Interesting quantitative measures

M1: **probability to successfully read the data**

M2: **99% time-quantile for successful reading**

Interesting quantitative measures

M1: probability to successfully read the data

M2: 99% time-quantile for successful reading

Interesting quantitative measures

M1: **probability to successfully read the data**

M2: **99% time-quantile for successful reading**

M3: fraction of time in which all replicas are damaged

M4: average time for repairing a damaged replica

M5: 99% time-quantile for repairing a damaged replica within time t

M6: probability to write at least c consistent replica within one write cycle

Interesting quantitative measures

M1: **probability to successfully read the data**

M2: **99% time-quantile for successful reading**

M3: fraction of time in which all replicas are damaged

M4: average time for repairing a damaged replica

M5: 99% time-quantile for repairing a damaged replica within time t

M6: probability to write at least c consistent replica within one write cycle

... on the long run ...

Long-run behavior

Steady-state distribution

Function $\theta : S \to [0,1]$ with

$$\theta(s) \stackrel{\text{def}}{=} \lim_{t \to \infty} \theta(s,t) \text{ with}$$

$\theta(s,t)$ the probability for being in state s at time $t \in \mathbb{R}_{\geq 0}$.

Long-run behavior

Steady-state distribution

Function $\theta : S \rightarrow [0,1]$ with

$$\theta(s) \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} \theta(s,t) \text{ with}$$

$\theta(s,t)$ the probability for being in state s at time $t \in \mathbb{R}_{\geq 0}$.

Important:

θ is well-defined distribution on S for finite CTMCs.

Long-run behavior

Steady-state distribution

Function $\theta : S \to [0, 1]$ with

$$\theta(s) \stackrel{\text{def}}{=} \lim_{t\to\infty} \theta(s, t) \text{ with}$$

$\theta(s, t)$ the probability for being in state s at time $t \in \mathbb{R}_{\geq 0}$.

Important:

θ is well-defined distribution on S for finite CTMCs.

Long-run probabilities

Let $\mathcal{M} = \langle S, Act, R, \mu\rangle$ be a CTMC. We refer to the probability measure obtained for the CTMC $\mathcal{M}_\theta = \langle S, Act, R, \theta\rangle$.

Conditional long-run behavior

Probability measure

Let $\mathcal{M} = \langle S, Act, R, \mu \rangle$ be a CTMC and $U \subseteq S$ be a set of states s.t. $\theta(U) > 0$. We refer to the probability measure obtained for the CTMC $\mathcal{M}_\theta^U = \mathcal{M}_\nu = \langle S, Act, R, \nu \rangle$

Conditional long-run behavior

Probability measure

Let $\mathcal{M} = \langle S, Act, R, \mu \rangle$ be a CTMC and $U \subseteq S$ be a set of states s.t. $\theta(U) > 0$. We refer to the probability measure obtained for the CTMC $\mathcal{M}_\theta^U = \mathcal{M}_\nu = \langle S, Act, R, \nu \rangle$, where

$$\nu(s) \;=\; \begin{cases} 0 & \text{if } s \in S \setminus U \\ \theta(s)/\theta(U) & \text{if } s \in U \end{cases}$$

Conditional long-run behavior

Probability measure

Let $\mathcal{M} = \langle S, Act, R, \mu \rangle$ be a CTMC and $U \subseteq S$ be a set of states s.t. $\theta(U) > 0$. We refer to the probability measure obtained for the CTMC $\mathcal{M}_\theta^U = \mathcal{M}_\nu = \langle S, Act, R, \nu \rangle$, where

$$\nu(s) \;=\; \begin{cases} 0 & \text{if } s \in S \setminus U \\ \theta(s)/\theta(U) & \text{if } s \in U \end{cases}$$

Conditional long-run queries

$\Pr\big(\Pi \,\big|\, U\big)$: conditional long-run probability

where Π is a measurable set of infinite paths, $U \subseteq S$ a set of states with $\theta(U) > 0$.

Conditional long-run behavior

Probability measure

Let $\mathcal{M} = \langle S, Act, R, \mu \rangle$ be a CTMC and $U \subseteq S$ be a set of states s.t. $\theta(U) > 0$. We refer to the probability measure obtained for the CTMC $\mathcal{M}_\theta^U = \mathcal{M}_\nu = \langle S, Act, R, \nu \rangle$, where

$$\nu(s) \;=\; \begin{cases} 0 & \text{if } s \in S \setminus U \\ \theta(s)/\theta(U) & \text{if } s \in U \end{cases}$$

Conditional long-run queries

$\mathrm{Pr}\big(\Pi \,\big|\, U\big)$: conditional long-run probability

$\mathrm{AccRew}\big(\lozenge T \,\big|\, U\big)$: conditional long-run accumulated reward

where Π is a measurable set of infinite paths, $U \subseteq S$ a set of states with $\theta(U) > 0$. We assume $\mathrm{Pr}\big(\lozenge T \,|\, U\big) = 1$.

Queries for interesting long run properties

Q1: probability to successfully read a replica

$$\Pr\big(\neg \mathit{error}^j \ \mathcal{U}\ \mathit{idle}^j \ \big|\ \mathit{reading\_started}^j_K\big)$$

Q2: time-quantile for successful reading within time bound t

$$\min\big\{t \ :\ p \leq \Pr\big(\neg \mathit{error}^j \ \mathcal{U}^{\leq t}\ \mathit{idle}^j \ \big|\ \mathit{reading\_started}^j_K\big)\big\}$$

Queries for interesting long run properties

Q3: fraction of time in which all replicas are damaged

$$\theta\big(\mathit{damaged}_1 \wedge \ldots \wedge \mathit{damaged}_K\big)$$

Queries for interesting long run properties

Q3: fraction of time in which all replicas are damaged

$$\theta\big(\mathit{damaged}_1 \wedge \ldots \wedge \mathit{damaged}_K\big)$$

Q4: average time for repairing a damaged replica

$$\mathrm{AccRew}\big(\Diamond\, \mathit{consistent}_k \;\big|\; \mathit{just\_damaged}_k\big)$$

Queries for interesting long run properties

Q3: fraction of time in which all replicas are damaged

$$\theta\big(\mathit{damaged}_1 \wedge \ldots \wedge \mathit{damaged}_K\big)$$

Q4: average time for repairing a damaged replica

$$\mathrm{AccRew}\big(\Diamond\, \mathit{consistent}_k \;\big|\; \mathit{just\_damaged}_k\big)$$

Q5: time-quantile for repairing a damaged replica within time t

$$\min\big\{t \;:\; p \leq \Pr\big(\Diamond^{\leq t}\, \mathit{consistent}_k \;\big|\; \mathit{just\_damaged}_k\big)\big\}$$

Queries for interesting long run properties

Q3: fraction of time in which all replicas are damaged

$$\theta\big(\mathit{damaged}_1 \wedge \ldots \wedge \mathit{damaged}_K\big)$$

Q4: average time for repairing a damaged replica

$$\mathrm{AccRew}\big(\Diamond\, \mathit{consistent}_k \;\big|\; \mathit{just\_damaged}_k\big)$$

Q5: time-quantile for repairing a damaged replica within time t

$$\min\big\{t \;:\; p \leq \Pr\big(\Diamond^{\leq t}\, \mathit{consistent}_k \;\big|\; \mathit{just\_damaged}_k\big)\big\}$$

Q6: probability to write at least c replica within one cycle

$$\Pr\big(\Pi_c \;\big|\; \mathit{writing\_started}_1^i\big)$$

Outline

Selected parameters and scenarios

Common parameters

	time	**rate**
write duration	2	$\lambda = 0.5$
read duration	1	$\delta = 1$
other	0.01	$\mu = \rho = \sigma = \nu = 100$

Selected parameters and scenarios

Common parameters

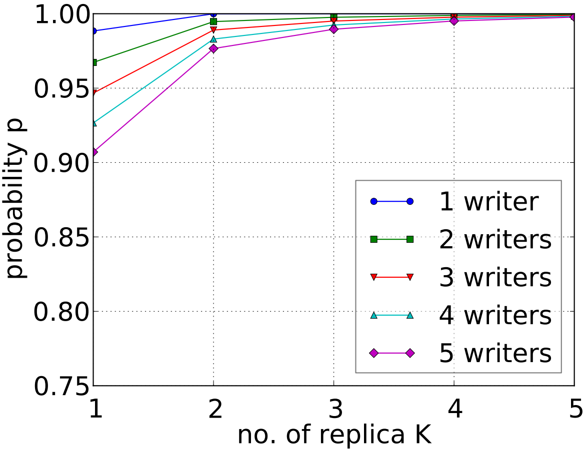
	time	rate
write duration	2	$\lambda = 0.5$
read duration	1	$\delta = 1$
other	0.01	$\mu = \rho = \sigma = \nu = 100$

Selected scenarios

	frequent reads moderate writes		moderate reads moderate writes	
	time	rate	time	rate
idle time (writer)	20	$\gamma = 0.05$	200	$\gamma = 0.005$
idle time (reader)	2	$\kappa = 0.5$	20	$\kappa = 0.05$

Results

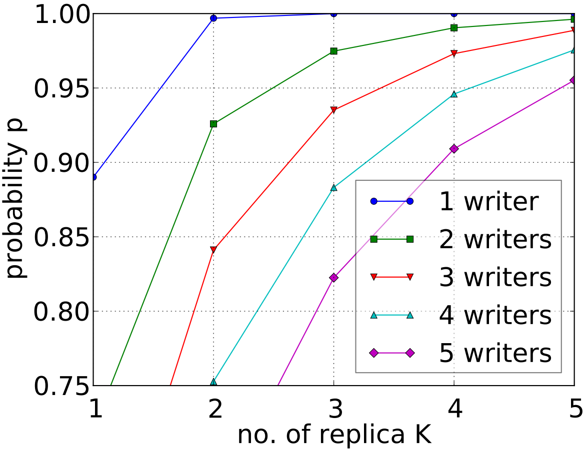
Q1: probability to successfully read the data
moderate reads, moderate writes



Results

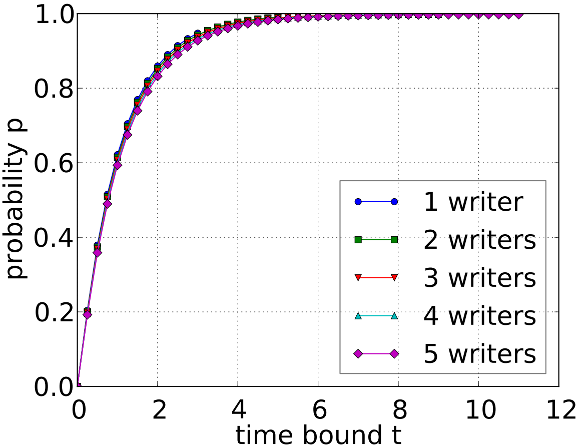
Q1: probability to successfully read the data

frequent reads, moderate writes



Results

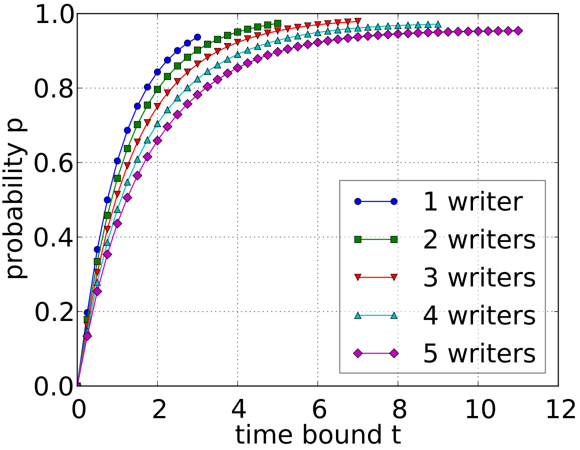
Q2: time-quantile for successful reading
moderate reads, moderate writes, 5 replicas



Results

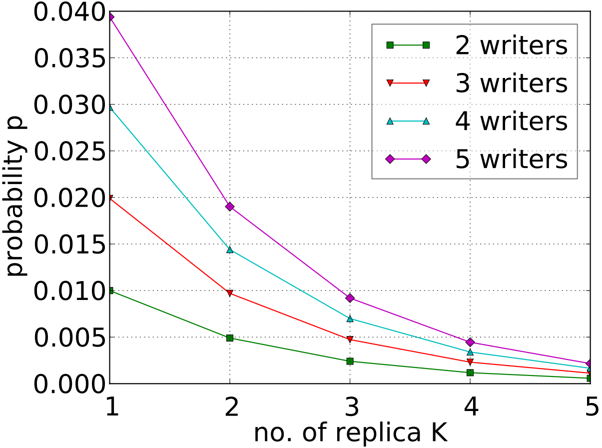
Q2: time-quantile for successful reading

frequent reads, moderate writes, 5 replicas



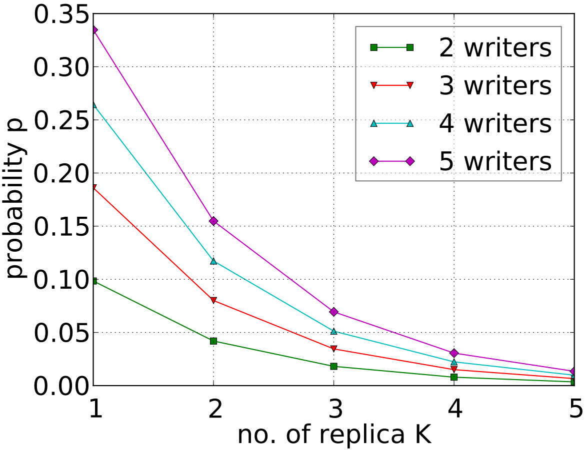
Results

Q3: time fraction in which all replicas are damaged
moderate reads, moderate writes



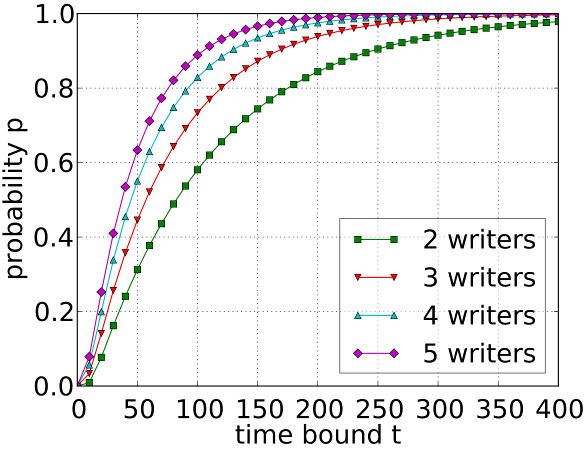
Results

Q3: time fraction in which all replicas are damaged
frequent reads, moderate writes



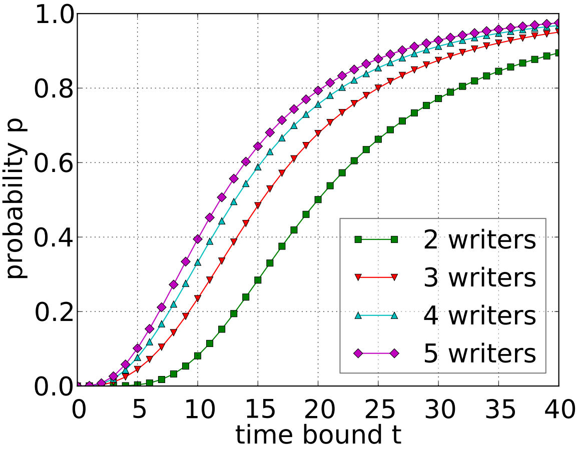
Results

Q5: time-quantile for repairing a damaged replica within time *t*
moderate reads, moderate writes, 5 replicas



Results

Q5: time-quantile for repairing a damaged replica within time *t*
frequent reads, moderate writes, 5 replicas



Outline

Conclusion and future work

Conclusion

- CTMC model for PWCS with multiple writers
- identification of quantitative measures for the evaluation of PWCS
- formalization of quantitative measures in terms of CSRL queries
- formal quantitative analysis of PWCS using PRISM

Conclusion and future work

Conclusion

- CTMC model for PWCS with multiple writers
- identification of quantitative measures for the evaluation of PWCS
- formalization of quantitative measures in terms of CSRL queries
- formal quantitative analysis of PWCS using PRISM

Future work

- comparative quantitative analysis with alternative protocols
- stronger object consistency in PWCS (e.g., multiple objects)
- other synchronization primitives (e.g., barriers)
- formal methods for quantile and (conditional) long run properties